PAC-Bayes Iterated Logarithm Bounds for Martingale Mixtures

Akshay Balsubramani
University of California, San Diego
abalsubr@cs.ucsd.edu

June 23, 2015

Abstract

We give tight concentration bounds for mixtures of martingales that are simultaneously uniform over (a) mixture distributions, in a PAC-Bayes sense; and (b) all finite times. These bounds are proved in terms of the martingale variance, extending classical Bernstein inequalities, and sharpening and simplifying prior work.

1 Introduction

The concentration behavior of a martingale M_t – a discrete-time stochastic process with conditionally stationary increments – is well-known to have many applications in modeling sequential processes and algorithms, and so it is of interest to analyze for applications in machine learning and statistics. It is a long-studied phenomenon [4]; despite their mighty generality, martingales exhibit essentially the same well-understood concentration behavior as simple random walks.

Even more powerful concentration results can be obtained by considering aggregates of many martingales. Though these too have long been studied asymptotically [8], their non-asymptotic study was only initiated by recent paper of Seldin et al. [6], which proves the sharpest known results on concentration of martingale mixtures, *uniformly* over the mixing distribution in a "PAC-Bayes" sense which is essentially optimal for such bounds [2]. This is motivated by and originally intended for applications in learning theory, as further discussed in that paper.

In this manuscript, we simplify, strengthen, and subsume the results of [6]. While that paper follows classical central-limit-theorem-type concentration results in focusing on an arbitrary fixed time, we instead leverage a recent method in Balsubramani [1] to achieve concentration that is uniform over finite times, extending the law of the iterated logarithm (LIL), with a rate at least as good as [6] and often far superior.

In short, our bounds on mixtures of martingales are uniform both over the mixing distribution in a PAC-Bayes sense, and over finite times, all simultaneously (and optimally). This has no precedent in the literature.

1.1 Preliminaries

To formalize this, consider a a set of discrete-time stochastic processes $\{M_t(h)\}_{h\in\mathcal{H}}$, where $\mathcal{H}$ is possibly uncountable, in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$. [1] Define the corresponding difference sequences $\xi_t(h) = M_t(h) - M_{t-1}(h)$ (which are $\mathcal{F}_t$-measurable for any h, t) and conditional variance processes $V_t(h) = \sum_{i=1}^t \mathbb{E}\left[\xi_i^2(h) \mid \mathcal{F}_{i-1}\right]$. The mean of the processes at time t w.r.t. any distribution ρ over $\mathcal{H}$ (any $\rho \in \Delta(\mathcal{H})$, as we write) is denoted by $\mathbb{E}_\rho[M_t] := \mathbb{E}_{h\sim\rho}[M_t(h)]$, with $\mathbb{E}_\rho[V_t]$ being defined similarly. For brevity, we drop the index h when it is clear from context.

Also recall the following standard definitions from martingale theory. For any $h \in \mathcal{H}$, a martingale M_t (resp. supermartingale, submartingale) has $\mathbb{E}[\xi_t \mid \mathcal{F}_{t-1}] = 0$ (resp. ≤ 0, ≥ 0) for all t. A stopping time τ is a function on Ω such that $\{\tau \leq t\} \in \mathcal{F}_t$ for all t; notably, τ can be infinite with positive probability ([4]).

It is now possible to state our main result.

[1] As in [1], we just consider discrete time for convenience; the core results and proofs in this paper extend to continuous time as well, as well as other generalizations discussed in that paper.

Theorem 1 (PAC-Bayes Martingale Bernstein LIL Concentration)**.** *Let $\{M_t(h)\}_{h\in\mathcal{H}}$ be a set of martingales and fix a distribution $\pi \in \Delta(\mathcal{H})$ and a $\delta < 1$. Suppose the difference sequence is uniformly bounded: $|M_t(h) - M_{t-1}(h)| \leq e^2$ for all $t \geq 1$ and $h \in \mathcal{H}$ w.p. 1. Then with probability $\geq 1-\delta$, the following is true for all $\rho \in \Delta(\mathcal{H})$.*

Suppose $\tau_0(\rho) = \min\left\{ s : 2(e-2)\mathbb{E}_\rho[V_s] \geq \frac{2}{\lambda_0^2}\left(\ln\left(\frac{4}{\delta}\right) + KL(\rho \mid\mid \pi)\right)\right\}$. For all $t \geq \tau_0(\rho)$ simultaneously, $|\mathbb{E}_\rho[M_t]| \leq \frac{2(e-2)}{e^2\left(1+\sqrt{1/3}\right)}\mathbb{E}_\rho[V_t]$ and

$$|\mathbb{E}_\rho[M_t]| \leq \sqrt{6(e-2)\mathbb{E}_\rho[V_t]\left(2\ln\ln\left(\frac{3(e-2)\mathbb{E}_\rho[V_t]}{|\mathbb{E}_\rho[M_t]|}\right) + \ln\left(\frac{2}{\delta}\right) + KL(\rho \mid\mid \pi)\right)}$$

(This bound is implicit in $|\mathbb{E}_\rho[M_t]|$, but notice that either $|\mathbb{E}_\rho[M_t]| \leq 1$ or the iterated-logarithm term can be simply treated as $2\ln\ln(3(e-2)\mathbb{E}_\rho[V_t])$, making the bound explicit.)

As we mentioned, Theorem 1 is uniform over ρ and t, allowing us to track mixtures of martingales tightly as they evolve. The martingales indexed by $\mathcal{H}$ can be highly dependent on each other, as they share a probability space. For instance, $M_{t_0}(h_0)$ can depend *in arbitrary fashion* on $\{M_{t\leq t_0}(h)\}_{h\neq h_0}$; it is only required to satisfy the martingale property for h_0, as further discussed in [6]. So these inequalities have found use in analyzing sequentially dependent processes such as those in reinforcement learning and bandit algorithms, where the choice of prior π can be tailored to the problem [5] and the posterior can be updated based on information learned up to that time.

The method of proof is essentially that used in [1]. Our main observation in this manuscript is that this proof technique is, in a technical sense, quite complementary to the fundamental method used in all PAC-Bayes analysis [2]. This allows us to prove our results in a sharper and more streamlined way than previous work [6].

1.2 Discussion

Let us elaborate on these claims. Theorem 1 can be compared directly to the following PAC-Bayes Bernstein bound from [6] that holds for a fixed time:

Theorem 2 (Restatement[2] of Thm. 8 from [6])**.** *Fix any t. For ρ such that $KL(\rho \mid\mid \pi)$ is sufficiently low compared to $\mathbb{E}_\rho[V_t]$,*

$$|\mathbb{E}_\rho[M_t]| \leq \sqrt{(1+e)^2(e-2)\mathbb{E}_\rho[V_t]\left(\ln\ln\left(\frac{(e-2)t}{\ln(2/\delta)}\right) + \ln\left(\frac{2}{\delta}\right) + KL(\rho \mid\mid \pi)\right)}$$

This bound is inferior to Theorem 1 in two significant ways: it holds for a fixed time rather than uniformly over finite times, and has an iterated-logarithm term of $\ln\ln t$ rather than $\ln\ln V_t$. The latter is a very significant difference when $V_t \ll t$, which is precisely when Bernstein inequalities would be preferred to more basic inequalities like Chernoff bounds.

Put differently, our non-asymptotic result, like those of Balsubramani [1], adapts correctly to the scale of the problem. We say "correctly" because Theorem 1 is optimal by the (asymptotic) martingale LIL, e.g. the seminal result of Stout [7]; this is true non-asymptotically too, by the main anti-concentration bound of [1]. All these optimality results are for a single martingale, but suffice for the PAC-Bayes case as well; and the additive $KL(\rho \mid\mid \pi)$ cost of uniformity over ρ is unimprovable in general also, by standard PAC-Bayes arguments.

1.2.1 Proof Overview

Our method follows that of Balsubramani [1], departing from the more traditionally learning-theoretic techniques used in [6]. We embark on the proof by introducing a standard exponential supermartingale construction that holds for any of the martingales $\{M_t(h)\}_{h\in\mathcal{H}}$.

[2] The actual statement of the theorem in [6], though not significantly different, is more complicated because of a few inconvenient artifacts of that paper's more complicated analysis, none of which arise in our analysis.

Lemma 3. *Suppose $|\xi_t| \leq e^2$ a.s. for all t. Then for any $h \in \mathcal{H}$, the process $X_t^\lambda(h) := \exp\left(\lambda M_t(h) - \lambda^2(e-2)V_t(h)\right)$ is a supermartingale in t for any $\lambda \in \left[-\frac{1}{e^2}, \frac{1}{e^2}\right]$.*

Proof. It can be checked that $e^x \leq 1 + x + (e-2)x^2$ for $x \leq 1$. Then for any $\lambda \in \left[-\frac{1}{e^2}, \frac{1}{e^2}\right]$ and $t \geq 1$,

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\xi_t\right) \mid \mathcal{F}_{t-1}\right] &\leq 1 + \lambda\mathbb{E}\left[\xi_t \mid \mathcal{F}_{t-1}\right] + \lambda^2(e-2)\mathbb{E}\left[\xi_t^2 \mid \mathcal{F}_{t-1}\right] \\ &= 1 + \lambda^2(e-2)\mathbb{E}\left[\xi_t^2 \mid \mathcal{F}_{t-1}\right] \leq \exp\left(\lambda^2(e-2)\mathbb{E}\left[\xi_t^2 \mid \mathcal{F}_{t-1}\right]\right)\end{aligned}$$

using the martingale property on $\mathbb{E}\left[\xi_t \mid \mathcal{F}_{t-1}\right]$. Therefore, $\mathbb{E}\left[\exp\left(\lambda\xi_t - \lambda^2(e-2)\mathbb{E}\left[\xi_t^2 \mid \mathcal{F}_{t-1}\right]\right) \mid \mathcal{F}_{t-1}\right] \leq 1$, so $\mathbb{E}\left[X_t^\lambda \mid \mathcal{F}_{t-1}\right] \leq X_{t-1}^\lambda$. $\square$

The classical martingale Bernstein inequality for a given h and fixed time t can be proved by using Markov's inequality with $\mathbb{E}\left[X_t^{\lambda^*}\right]$, where $\lambda^* \propto \frac{|M_t|}{V_t}$ is tuned for the tightest bound, and can be thought of as setting the relative scale of variation being measured.

The proof technique of this paper and its advantages over previous work are best explained by examining how to set the scale parameter λ.

1.2.2 Choosing the Scale Parameter

On a high level, the main idea of Balsubramani [1] is to average over a random choice of the scale parameter λ in the supermartingale X_t^λ. This allows manipulation of a stopped version of X_t^λ, i.e. X_τ^λ for a particular stopping time τ. So the averaging technique in [1] can be thought of as using "many values of λ at once," which is necessary when dealing with the stopped process because τ is random, and so is $\frac{|M_\tau|}{V_\tau}$.

All existing PAC-Bayes analyses achieve uniformity in ρ through the Donsker-Varadhan variational characterization of the KL divergence:

Lemma 4 (Donsker-Varadhan Lemma ([3])). *Suppose ρ and π are probability distributions over $\mathcal{H}$, and let $f(\cdot)\colon \mathcal{H} \mapsto \mathbb{R}$ be a measurable function. Then*

$$\mathbb{E}_\rho\left[f(h)\right] \leq KL(\rho \mid\mid \pi) + \ln\left(\mathbb{E}_\pi\left[e^{f(h)}\right]\right)$$

This introduces a $KL(\rho \mid\mid \pi)$ term into the bounding of X_t^λ. However, the optimum λ^* is then dependent on the unknown ρ. The solution adopted by existing PAC-Bayes martingale bounds ([6] and variants) is again to use "many values of λ at once." In prior work, this is done explicitly by taking a union bound over a grid of carefully chosen λs.

Our main technical contribution is to recognize the similarity between these two problems, and to use the (tight) stochastic choice of λ in [1] as a solution to both problems at once, achieving the optimal bound of Theorem 1.

2 Proof of Theorem 1

We now give the complete proof of Theorem 1, following the presentation of [1] closely.

For the rest of this section, define $U_t := 2(e-2)V_t$, $k := \frac{1}{3}$, and $\lambda_0 := \frac{1}{e^2\left(1+\sqrt{k}\right)}$. As in [1], our proof invokes the Optional Stopping Theorem from martingale theory, in particular a version for nonnegative supermartingales that neatly exploits their favorable convergence properties:

Theorem 5 (Optional Stopping for Nonnegative Supermartingales ([4], Theorem 5.7.6)). *If M_t is a nonnegative supermartingale and τ a (possibly infinite) stopping time, $\mathbb{E}\left[M_\tau\right] \leq \mathbb{E}\left[M_0\right]$.*

We also use the exponential supermartingale construction of Lemma 3, which we assume to hold for $M_t(h)$ $\forall h \in \mathcal{H}$ since they are all martingales whose concentration we require.

Our goal is to control $\mathbb{E}_\rho\left[M_t\right]$ in terms of $\mathbb{E}_\rho\left[U_t\right]$, so it is tempting to try to show that $e^{\lambda\mathbb{E}_\rho[M_t] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_t]}$ is an exponential supermartingale. However, this is not generally true; and even if it were, would only control $\mathbb{E}\left[e^{\lambda\mathbb{E}_\rho[M_\tau] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_\tau]}\right]$ for a fixed ρ, not in a PAC-Bayes sense.

We instead achieve uniform control over ρ by using the Donsker-Varadhan lemma to mediate the Optional Stopping Theorem every time it is used in Balsubramani's proof [1] of the single-martingale case. This is fully captured in the following key result, a powerful extension of a standard moment-generating function bound that is uniform in ρ and has enough freedom (an arbitrary stopping time τ) to be converted into a time-uniform bound.

Lemma 6. *Choose any probability distribution π over $\mathcal{H}$. Then for any stopping time τ and $\lambda \in \left[-\frac{1}{e^2}, \frac{1}{e^2}\right]$, simultaneously for all distributions ρ over $\mathcal{H}$,*

$$\mathbb{E}\left[e^{\lambda \mathbb{E}_\rho[M_\tau] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_\tau]}\right] \leq e^{KL(\rho||\pi)}$$

Proof. Using Lemma 4 with the function $f(h) = \lambda M_\tau(h) - \frac{\lambda^2}{2} U_\tau(h)$, and exponentiating both sides, we have for all posterior distributions $\rho \in \Delta(\mathcal{H})$ that

$$e^{\lambda \mathbb{E}_\rho[M_\tau] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_\tau]} \leq e^{KL(\rho||\pi)} \mathbb{E}_\pi\left[e^{\lambda M_\tau - \frac{\lambda^2}{2} U_\tau}\right] \tag{1}$$

Therefore, $\mathbb{E}\left[e^{\lambda \mathbb{E}_\rho[M_\tau] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_\tau]}\right] \overset{(a)}{\leq} e^{KL(\rho||\pi)} \mathbb{E}_\pi\left[\mathbb{E}\left[e^{\lambda M_\tau - \frac{\lambda^2}{2} U_\tau}\right]\right] \overset{(b)}{\leq} e^{KL(\rho||\pi)}$ where (a) is from (1) and Tonelli's Theorem, and (b) is by Lemma 3 and Optional Stopping (Thm. 5). $\square$

Just as a bound on the moment-generating function of a random variable is the key to proving tight Hoeffding and Bernstein concentration of that variable, Lemma 6 is, exactly analogously, the key tool used to prove Theorem 1.

2.1 A PAC-Bayes Uniform Law of Large Numbers

First, we define the stopping time $\tau_0(\rho) := \min\left\{t : \mathbb{E}_\rho[U_t] \geq \frac{2}{\lambda_0^2}\left(\ln\left(\frac{2}{\delta}\right) + KL(\rho \mid\mid \pi)\right)\right\}$ and the following "good" event:

$$B_\delta = \left\{\omega \in \Omega : \forall \rho \in \Delta(\mathcal{H}), \quad \frac{|\mathbb{E}_\rho[M_t]|}{\mathbb{E}_\rho[U_t]} \leq \lambda_0 \qquad \forall t \geq \tau_0(\rho)\right\} \tag{2}$$

Our first result introduces the reader to our main proof technique; it is a generalization of the law of large numbers (LLN) to our PAC-Bayes martingale setting.

Theorem 7. *Fix any $\delta > 0$. With probability $\geq 1 - \delta$, the following is true for all ρ over $\mathcal{H}$: for all $t \geq \tau_0(\rho)$,*

$$\frac{|\mathbb{E}_\rho[M_t]|}{\mathbb{E}_\rho[U_t]} \leq \lambda_0$$

To prove this, we first manipulate Lemma 6 so that it is in terms of $|\mathbb{E}_\rho[M_\tau]|$.

Lemma 8. *Choose any prior $\pi \in \Delta(\mathcal{H})$. For any stopping time τ and all distributions ρ over $\mathcal{H}$,*

$$\mathbb{E}\left[\exp\left(\lambda_0 |\mathbb{E}_\rho[M_\tau]| - \frac{\lambda_0^2}{2}\mathbb{E}_\rho[U_\tau]\right)\right] \leq 2e^{KL(\rho||\pi)}$$

Proof. Lemma 6 describes the behavior of the process $\chi_t^\lambda = e^{\lambda \mathbb{E}_\rho[M_t] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_t]}$. Define Y_t to be the mean of χ_t^λ with λ being set *stochastically*: $\lambda \in \{-\lambda_0, \lambda_0\}$ with probability $\frac{1}{2}$ each. After marginalizing over λ, the resulting process is

$$Y_t = \frac{1}{2}\exp\left(\lambda_0 \mathbb{E}_\rho[M_t] - \frac{\lambda_0^2}{2}\mathbb{E}_\rho[U_t]\right) + \frac{1}{2}\exp\left(-\lambda_0 \mathbb{E}_\rho[M_t] - \frac{\lambda_0^2}{2}\mathbb{E}_\rho[U_t]\right) \geq \frac{1}{2}\exp\left(\lambda_0 |\mathbb{E}_\rho[M_t]| - \frac{\lambda_0^2}{2}\mathbb{E}_\rho[U_t]\right) \tag{3}$$

Now take τ to be any stopping time as in the lemma statement. Then $\mathbb{E}\left[\exp\left(\lambda_0 \mathbb{E}_\rho[M_\tau] - \frac{\lambda_0^2}{2}\mathbb{E}_\rho[U_\tau]\right)\right] \leq e^{KL(\rho||\pi)}$, by Lemma 6. Similarly, $\mathbb{E}\left[X_\tau^{\lambda=-\lambda_0}\right] \leq e^{KL(\rho||\pi)}$.

So $\mathbb{E}[Y_\tau] = \frac{1}{2}\left(\mathbb{E}\left[X_\tau^{\lambda=-\lambda_0}\right] + \mathbb{E}\left[X_\tau^{\lambda=\lambda_0}\right]\right) \leq e^{KL(\rho||\pi)}$. Combining this with (3) gives the result. $\square$

A particular setting of τ extracts the desired uniform LLN bound from Lemma 8.

Proof of Theorem 7. Define the stopping time

$$\tau = \min \left\{ t : \exists \rho \in \Delta(\mathcal{H}) \text{ s.t. } t \geq \tau_0(\rho) \text{ and } \frac{|\mathbb{E}_\rho [M_t]|}{\mathbb{E}_\rho [U_t]} > \lambda_0 \right\}$$

Then it suffices to prove that $P(\tau < \infty) \leq \delta$.

On the event $\{\tau < \infty\}$, we have for some ρ that $\frac{|\mathbb{E}_\rho[M_\tau]|}{\mathbb{E}_\rho[U_\tau]} > \lambda_0$ by definition of τ. Therefore, using Lemma 8, we can conclude that for this ρ,

$$2e^{KL(\rho||\pi)} \geq \mathbb{E}\left[\exp\left(\lambda_0 |\mathbb{E}_\rho [M_\tau]| - \frac{\lambda_0^2}{2}\mathbb{E}_\rho [U_\tau]\right) \mid \tau < \infty\right] P(\tau < \infty) \overset{(b)}{\geq} \frac{2}{\delta} e^{KL(\rho||\pi)} P(\tau < \infty)$$

where (b) uses $\mathbb{E}_\rho [U_\tau] \geq \mathbb{E}_\rho [U_{\tau_0}] \geq \frac{2}{\lambda_0^2} \ln\left(\frac{2}{\delta} e^{KL(\rho||\pi)}\right)$. Therefore, $P(\tau < \infty) \leq \delta$, as desired. □

2.2 Proof of Theorem 1

For any event $E \subseteq \Omega$ of nonzero measure, let $\mathbb{E}_E [\cdot]$ denote the expectation restricted to E, i.e. $\mathbb{E}_E [f] = \frac{1}{P(E)} \int_E f(\omega) P(d\omega)$ for a measurable function f on Ω. Similarly, dub the associated measure P_E, where for any event $\Xi \subseteq \Omega$ we have $P_E(\Xi) = \frac{P(E \cap \Xi)}{P(E)}$.

Theorem 7 shows that $P(B_\delta) \geq 1 - \delta$. It is consequently easy to observe that the corresponding restricted outcome space can still be used to control expectations.

Proposition 9. *For any nonnegative r.v. X, we have $\mathbb{E}_{B_\delta} [X] \leq \frac{1}{1-\delta} \mathbb{E} [X]$.*

Proof. Using Thm. 7, $\mathbb{E} [X] = \mathbb{E}_{B_\delta} [X] P(B_\delta) + \mathbb{E}_{B_\delta^c} [X] P(B_\delta^c) \geq \mathbb{E}_{B_\delta} [X] (1 - \delta)$. □

Just as in [1], the idea of the main proof is to choose λ stochastically from a probability space $(\Omega_\lambda, \mathcal{F}_\lambda, P_\lambda)$ such that $P_\lambda(d\lambda) = \dfrac{d\lambda}{|\lambda| \left(\log \frac{1}{|\lambda|}\right)^2}$ on $\lambda \in [-e^{-2}, e^{-2}] \setminus \{0\}$. The parameter λ is chosen independently of $\xi_1, \xi_2, \ldots$, so that X_t^λ is defined on the product space. Write $\mathbb{E}^\lambda [\cdot]$ to denote the expectation with respect to $(\Omega_\lambda, \mathcal{F}_\lambda, P_\lambda)$.

To be consistent with previous notation, we continue to write $\mathbb{E} [\cdot]$ to denote the expectation w.r.t. the original probability space $(\Omega, \mathcal{F}, P)$. As mentioned earlier, we use subscripts for expectations conditioned on events in this space, e.g. $\mathbb{E}_{A_\delta} [X]$. (As an example, $\mathbb{E}_\Omega [\cdot] = \mathbb{E} [\cdot]$.)

Before going on to prove the main theorem, we need one more result that controls the effect of averaging over λ as above.

Lemma 10. *For all $\rho \in \Delta(\mathcal{H})$ and any δ, the following is true: for any stopping time $\tau \geq \tau_0(\rho)$,*

$$\mathbb{E}_{B_\delta} \left[\mathbb{E}^\lambda \left[e^{\lambda \mathbb{E}_\rho [M_\tau] - \frac{\lambda^2}{2} \mathbb{E}_\rho [U_\tau]}\right]\right] \geq \mathbb{E}_{B_\delta} \left[\frac{2 \exp\left(\frac{\mathbb{E}_\rho [M_\tau]^2}{2\mathbb{E}_\rho [U_\tau]} (1 - k)\right)}{\ln^2 \left(\frac{\mathbb{E}_\rho [U_\tau]}{(1-\sqrt{k}) |\mathbb{E}_\rho [M_\tau]|}\right)}\right]$$

Lemma 10 is precisely analogous to Lemma 13 in [1] and proved using exactly the same calculations, so its proof is omitted here. Now we can prove Theorem 1.

Proof of Theorem 1. The proof follows precisely the same method as that of Balsubramani [1], but with a more nuanced setting of the stopping time τ. To define it, we first for convenience define the deterministic function

$$\zeta_t(\rho) = \sqrt{\frac{2\mathbb{E}_\rho [U_t]}{1 - k} \ln \left(\frac{2 \ln^2 \left(\frac{\mathbb{E}_\rho [U_t]}{(1-\sqrt{k}) |\mathbb{E}_\rho [M_t]|}\right) e^{KL(\rho||\pi)}}{\delta}\right)}$$

Now we define the stopping time

$$\tau = \min \Bigg\{ t:\ \exists \rho \in \Delta(\mathcal{H})\ \text{ s.t. }\ t \geq \tau_0(\rho) \quad \text{and}$$
$$\left[|\mathbb{E}_\rho[M_t]| > \lambda_0 \mathbb{E}_\rho[U_t] \quad \text{or} \quad (|\mathbb{E}_\rho[M_t]| \leq \lambda_0 \mathbb{E}_\rho[U_t] \ \text{ and } \ |\mathbb{E}_\rho[M_t]| > \zeta_t(\rho)) \right] \Bigg\}$$

The rest of the proof shows that $P(\tau = \infty) \geq 1 - \delta$, and involves nearly identical calculations to the main proof of [1] (with $\mathbb{E}_\rho[M_t], \mathbb{E}_\rho[U_t], B_\delta$ replacing what that paper writes as M_t, U_t, A_δ).

It suffices to prove that $P(\tau = \infty) \geq 1 - \delta$. On the event $\left\{ \{\tau < \infty\} \cap B_{\delta/2} \right\}$, by definition there exists a ρ s.t.

$$|\mathbb{E}_\rho[M_\tau]| > \zeta_t(\rho) = \sqrt{\frac{2\mathbb{E}_\rho[U_\tau]}{1-k} \ln \left(\frac{2\ln^2\left(\frac{\mathbb{E}_\rho[U_\tau]}{(1-\sqrt{k})|\mathbb{E}_\rho[M_\tau]|} \right)}{\delta} e^{KL(\rho||\pi)} \right)}$$

$$\text{I.e. a } \rho \text{ such that} \qquad \frac{2\exp\left(\frac{\mathbb{E}_\rho[M_\tau]^2}{2\mathbb{E}_\rho[U_\tau]}(1-k) \right)}{\ln^2\left(\frac{\mathbb{E}_\rho[U_\tau]}{(1-\sqrt{k})|\mathbb{E}_\rho[M_\tau]|} \right)} > \frac{4}{\delta} e^{KL(\rho||\pi)} \tag{4}$$

Consider this ρ. Using the nonnegativity of $\frac{2\exp\left(\frac{\mathbb{E}_\rho[M_t]^2}{2\mathbb{E}_\rho[U_t]}(1-k) \right)}{\ln^2\left(\frac{\mathbb{E}_\rho[U_t]}{(1-\sqrt{k})|\mathbb{E}_\rho[M_t]|} \right)}$ on $B_{\delta/2}$ and letting $Z_\tau^\lambda := e^{\lambda \mathbb{E}_\rho[M_\tau] - \frac{\lambda^2}{2}\mathbb{E}_\rho[U_\tau]}$,

$$\begin{aligned}
2e^{KL(\rho||\pi)} &\geq \frac{e^{KL(\rho||\pi)}}{1-\frac{\delta}{2}} \overset{(a)}{\geq} \frac{\mathbb{E}^\lambda\left[\mathbb{E}\left[Z_\tau^\lambda\right]\right]}{1-\frac{\delta}{2}} \overset{(b)}{\geq} \mathbb{E}^\lambda\left[\mathbb{E}_{B_{\delta/2}}\left[Z_\tau^\lambda\right]\right] \overset{(c)}{=} \mathbb{E}_{B_{\delta/2}}\left[\mathbb{E}^\lambda\left[Z_\tau^\lambda\right]\right] \\
&\overset{(d)}{\geq} \mathbb{E}_{B_{\delta/2}}\left[\frac{2\exp\left(\frac{\mathbb{E}_\rho[M_\tau]^2}{2\mathbb{E}_\rho[U_\tau]}(1-k) \right)}{\ln^2\left(\frac{\mathbb{E}_\rho[U_\tau]}{(1-\sqrt{k})|\mathbb{E}_\rho[M_\tau]|} \right)} \right] \geq \mathbb{E}_{B_{\delta/2}}\left[\frac{2\exp\left(\frac{\mathbb{E}_\rho[M_\tau]^2}{2\mathbb{E}_\rho[U_\tau]}(1-k) \right)}{\ln^2\left(\frac{\mathbb{E}_\rho[U_\tau]}{(1-\sqrt{k})|\mathbb{E}_\rho[M_\tau]|} \right)} \mid \tau < \infty \right] P_{B_{\delta/2}}(\tau < \infty) \\
&\overset{(e)}{>} \frac{4}{\delta} e^{KL(\rho||\pi)} P_{B_{\delta/2}}(\tau < \infty)
\end{aligned}$$

where (a) is by Lemma 6, (b) is by Prop. 9, (c) is by Tonelli's Theorem, (d) is by Lemma 10, and (e) is by (4).

After simplification, this gives

$$P_{B_{\delta/2}}(\tau < \infty) \leq \delta/2 \implies P_{B_{\delta/2}}(\tau = \infty) \geq 1 - \frac{\delta}{2} \tag{5}$$

and using Theorem 7 – that $P(B_{\delta/2}) \geq 1 - \frac{\delta}{2}$ – concludes the proof. □

References

[1] Akshay Balsubramani. Sharp uniform martingale concentration bounds. *arXiv preprint arXiv:1405.2639*, 2015.

[2] Arindam Banerjee. On bayesian bounds. In *Proceedings of the 23rd international conference on Machine learning*, pages 81–88. ACM, 2006.

[3] M. D. Donsker and S. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time. iii. *Communications on Pure and Applied Mathematics*, 28:389–461, 1976.

[4] Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4th edition, 2010.

[5] Yevgeny Seldin, Peter Auer, John S Shawe-taylor, Ronald Ortner, and François Laviolette. Pac-bayesian analysis of contextual bandits. In *Advances in Neural Information Processing Systems*, pages 1683–1691, 2011.

[6] Yevgeny Seldin, François Laviolette, Nicolo Cesa-Bianchi, John Shawe-Taylor, and Peter Auer. Pac-bayesian inequalities for martingales. *Information Theory, IEEE Transactions on*, 58(12):7086–7093, 2012.

[7] William F. Stout. A martingale analogue of kolmogorov's law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete*, 15:279–290, 1970.

[8] R. J. Tomkins. Iterated logarithm results for weighted averages of martingale difference sequences. *Ann. Probab.*, 3(2):307–314, 04 1975.